are allowed to circulate freely, that is, without “tags” pointing to herself as their source, in Katerina Ivanovna’s mind they produce inferences that can corrupt the already existing stores of knowledge. After all, Katerina Ivanovna’s late father had been a socially prominent figure, and Petr Petrovich could have been, in principle, welcomed in his house, if the two men had ever had a chance to meet. What happens here is that Katerina Ivanovna’s original memory of her father’s house is now corrupted by the conviction that Petr Petrovich used to be a frequent guest there. (Compare it to the hypothetical situation above, in which the information that it is raining gold, when assimilated without a source-specifying tag, such as, “It was Eve who told me,” begins to impact our other knowledge stores and results in harmful behavior, such as canceling classes, quitting the job, maxing out on credit cards, etc.).

EVERYDAY FAILURES OF SOURCE-MONITORING

Of course, it is not just the hapless Katerina Ivanovna who invents stories about the state of affairs in the world and begins to act upon them as if they were real. We all do it. In many cases, such self-deception is quite beneficial—as one of the more level-headed (or just differently insane) characters from Crime and Punishment observes, “Best lives he who dupes himself the best” (502). But generally, especially if we consider the closely related issue of personal memories, it makes sense to think of our partial failures to keep track of some of the sources of our representations as part of the normal functioning of the metarepresenting brain. When I say “normal,” I mean to contrast it both with the sustained, pathological pattern of such failures typical for schizophrenic patients and with the deliberately planned and carefully highlighted instances of such failures in the works of fiction.

I was reminded some time ago about everyday failures of our source-monitoring—failures that we do not even register consciously unless pressed by circumstances—while reading the account of Martha Stewart’s trial in The New Yorker (Stewart had been accused of insider trading and subsequent lying to federal agents). The author, Jeffrey Toobin, refers to a “curious” testimony by one of Stewart’s close friends, Mariana Pasternak, who, at one point, could not identify the source of one of her memories:
Pasternak's appearance ended on a curious note. In her direct testimony, she said that, in another conversation in Mexico, Stewart had commented about [the tip of her broker who had advised her to sell her stocks in the biotech company ImClone]: ‘Isn't it nice to have brokers who tell you those things?’ But under [the defense lawyer's] cross-examination, she said, ‘I do not know if that statement was made by Martha or just was a thought in my mind’—a concession so dramatic that it brought a gasp from the spectators. But then, when the prosecution questioned her again, Pasternak said her ‘best belief’ was that Stewart said it. (70)

I suspect that the main reason Pasternak's concession “brought a gasp from the spectators” is the charged atmosphere of the courtroom and the specifics of this particular case, in which so much hinged on reconstructing who said exactly what and exactly when. Had any of the “gasping” spectators been asked to trace the exact sources of this or that representation of his, it is likely that he would feel just as uncertain about certain aspects of it as Pasternak did.¹

One may ask, then, why we should posit our metarepresentational ability as a special cognitive endowment when it seems that we are routinely unsure about the sources of our representations. The answer to this question applies equally well to the question of why we should posit our Theory of Mind as a very special cognitive adaptation when in fact we routinely misread, misinterpret, and misrepresent other people's states of mind. To adapt one of Ellen Spolsky’s insights, both the metarepresentational ability and the Theory of Mind are not “perfect” in some abstract, context-independent sense. Instead, they are “good enough”² for our everyday functioning: however imperfect and fallible, they still get us through yet another day of social interactions.

Thus, in the example above, the trial witness may have difficulties pinpointing the exact source of her personal memory, but even her apparent failure is thoroughly structured by her metarepresentational ability. That is, she knows that the representation, “Isn't it nice to have brokers who tell you those things?” does not simply describe the state of affairs but also expresses somebody’s opinion. Even if she strongly agrees with the truth of this sentiment, on some level it has still been processed in her mind with a tag limiting its source to two people, either herself or Martha Stewart. The potential for a misattribution or uncertainty (e.g., “Was it really me or Martha?”) falls within the same functional range as (to return to the example from Part I) our mistaken interpretation of tears of joy on our friend’s face as tears of grief. In the latter case, our range of readings is drastically and
productively limited to the domain of emotions; in the former case, Pasternak’s range of attribution is drastically and productively limited to two people (as opposed to, say, 150 other people of her acquaintance).

Though not “perfect” (in some rather abstract way), this is surely a “good enough” cognitive scenario, of the kind that we live with daily. Evolution, as Tooby and Cosmides frequently point out, did not have a crystal ball: the adaptations that contributed, with statistical reliability, to the survival of the human species for hundreds of thousands of years and thus became part of our permanent cognitive makeup profoundly structure our interaction with the world, but even when they function properly, at no point do they guarantee a smooth sailing through concrete complicated situations or the instinctive knowing of the exact origins of every aspect of our personal memories.

~4~

MONITORING FICTIONAL STATES OF MIND

However little we may know at this point about our metarepresentational ability, applying what we do know (or at least hypothesize strongly) to analysis of fiction results in the same embarrassment of riches as does the application of the Theory-of-Mind research. We start realizing that our capacity for storing representations under various degrees of advisement profoundly structures our interaction with literary texts, although, just as with the Theory of Mind, specific historical and cultural circumstances shape the specific forms that such interaction takes. Broadly speaking, whereas our Theory of Mind makes it possible for us to invest literary characters with a potential for a broad array of thoughts, desires, intentions, and feelings and then to look for textual cues that allow us to figure out their states of mind and thus predict their behavior, our metarepresentational ability allows us to discriminate among the streams of information coming at us via all this mind-reading. It allows us to assign differently weighed truth-values to representations originating from different sources (that is, characters, including the narrator) under specific circumstances. The ability to keep track of who thought, wanted, and felt what, and when they thought it, is crucial considering that the majority of our fictional narratives, from Homer’s The Iliad, Shikibu’s The Tale of Genji, and St. Augustine’s Confessions, to Tolstoy’s War and Peace and